DOCUMENT RESUME

ED 453 224 TM 032 575

AUTHOR Cole, Rebecca Pollard; MacIsaac, Dan; Cole, David M.
TITLE A Comparison of Paper-Based and Web-Based Testing.
PUB DATE 2001-04-13
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001). This research was partially supported by Northern Arizona University Organized Research Grant funds, the NAU Department of Physics and Astronomy, the National Science Foundation through the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT), and the U.S. Department of Education funded Arizona Teacher's Excellence Coalition (AzTEC).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Students; *Computer Assisted Testing; Higher Education; *Physics; *Test Format; Test Items; *World Wide Web
IDENTIFIERS *Paper and Pencil Tests

ABSTRACT

The purpose of this study (1,313 college student participants) was to examine the differences in paper-based and Web-based administrations of a commonly used assessment instrument, the Force Concept Inventory (FCI) (D. Hestenes, M. Wells, and G. Swackhamer, 1992). Results demonstrated no appreciable difference on FCI scores or FCI items based on the type of administration. Analyses demonstrated differences in FCI scores due to gender and time of administration (pre-test and post-test). However, none of these differences was influenced by the type of test administration (Web or paper). Similarly, FCI student scores were comparable with respect to test reliability. For individual FCI items, paper-based and Web-based comparisons were made by examining potential differences in item means and by examining potential differences in response patterns. Chi Squares demonstrated no differences in response patterns and t-Tests demonstrated no differences in item means between paper-based and Web-based administrations. In summary, the Web-based administration of the FCI appears to be as efficacious as the paper-based administration. Lessons learned from the implementation of Web-administered testing are also discussed. (Contains 2 figures, 4 tables, and 27 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 453 224

A Comparison of Paper-based and Web-based Testing

Rebecca Pollard Cole

Dan MacIsaac

David M. Cole

Northern Arizona University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
D. MacIsaac
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

TM032575

Reprint requests should be sent to Dan MacIsaac (danmac@nau.edu), Department of Physics and Astronomy, Northern Arizona University, Flagstaff, AZ, 86011-6010.

BEST COPY AVAILABLE

Abstract

On-line web-based technologies provide students with the opportunity to complete assessment instruments from personal computers with internet access. The purpose of this study was to examine the differences in paper-based and web-based administrations of a commonly used assessment instrument, the Force Concept Inventory (FCI). Results demonstrated no appreciable difference on FCI scores or FCI items based on the type of administration. Analyses demonstrated differences in FCI scores due to gender and time of administration (pre- and post-). However, none of these differences was influenced by the type of test administration (web or paper). Similarly, FCI student scores were comparable with respect to test reliability. For individual FCI items, paper-based and web-based comparisons were made by examining potential differences in item means and by examining potential differences in response patterns. Chi Squares demonstrated no differences in response patterns and t Tests demonstrated no differences in item means between paper-based and web-based administrations. In summary, the web-based administration of the Force Concept Inventory appears to be as efficacious as the paper-based administration. Lessons learned from the implementation of web-administered testing are also discussed.

A Comparison of Paper-based and Web-based Assessment

Since the late 1970's, science educators have been experimenting with the use of microcomputers for the conceptual and attitudinal assessment of their students (Arons, 1984, 1986; Bork, 1981; Waugh, 1985). Since the late 1980's, multiple-choice, machine scored, standardized instruments have been developed to assess the conceptual and attitudinal state of introductory physics students. The Force Concept Inventory (FCI), perhaps the best known of these standardized instruments, assesses student's conceptual knowledge of physics (see Hestenes, Wells & Swackhamer, 1992). Recently, Redish, Saul, and Steinberg (1998) developed the Maryland Physics Expectations Survey (MPEX), a standardized instrument which assesses the attitudinal state of physics students. Both the FCI and the MPEX are widely used in the physics education research (PER) community (Hake, 1998).

Although these instruments were initially used by experts for physics education research (PER) only, more generalized interests in program evaluation, curriculum development, justifying and guiding interventions in physics teaching practices and comparing student learning and attitudinal outcomes have led to widespread desires to use these instruments. Anticipating this interest, the FCI was published with the statement that "[the FCI] is included here for teachers to use in any way they see fit" (Hestenes, Wells & Swackhamer, 1992. p142). As one example of such use for program evaluation, the FCI was recently adopted as one of a suite of instruments to be used for the regular and routine assessment of student learning in the physics course sequences at Northern Arizona University (MacIsaac, 1999).

There are administrative burdens associated with standard use of these instruments. For instance, completion of one of these instruments requires approximately thirty minutes of class, laboratory or recitation time. Since these instruments are typically administered both pre- and post- instruction, each instrument could therefore consume up to an hour of

scarce and valuable instructional time. In addition, resources required to duplicate, administer, collect, collate, accurately code, score, record, and analyze the instrument data are sharply limited in many departments, strongly discouraging regular and routine paper-based administration of these instruments. Hake (1998) confirms that both the loss of instructional time and the administrative overhead may discourage the regular use of these instruments by many introductory physics instructors. Hence our interest in alternative, non-classroom administration of these instruments at NAU.

Web-based technologies provide students with an alternative to paper administration -- the opportunity to complete assessment instruments from personal computers via internet access (Titus, Martin & Beichner, 1998). Harvey and Mogey (1999) suggest economies of time, scale and student effort are possible by amortizing development of web-coding infrastructure over many semesters, eliminating the need for expensive optical scan forms, reusing instrument data for multiple reasons and establishing uniform assessment administrations for future, continuing student use in following courses. Danson (1999) suggests further advantages to web testing such as improved response accuracy by reducing input response errors such as skipped rows of optically marked bubbles and assuring statistical software interpretability by input checking and appropriately constrained input selection. Cann & Pawley (1999) note that web pages can reduce coding errors and write student-provided data directly to computer files that can themselves be used as input files for computerized statistical analysis, removing any further need to code data for computer input. Web-based administration of standardized instruments can even allow simultaneous collection of new kinds of data for improving the instruments themselves (such as question latency data -- the length of time required for responses).

Security is another issue: web-administered instruments appear to trade security for flexibility (Harvey & Mogey, 1999). Authentication (verifying the identity of the person completing an instrument) is difficult or impossible to ensure outside of a monitored

computer laboratory. Web test takers may be inappropriately collaborating with others, sharing questions with others, cheating or using reference materials.

Some student may also develop increased anxiety (Brosnan, 1999) associated with computer use that could lead to distorted data. Finally, all students may not have ready and appropriate access to computers and the web necessary to complete web administered instruments (Harvey & Mogey, 1999), which may become less of an issue for physics students as time progresses.

However, to be commensurate with the current collection of paper-administered FCI data, the equivalence or mapping for web-administered version of standardized physics instruments must be developed. As discussed by Brosnan (1999):

> The American Psychological Association's (1996) *Guidelines for Computer-based tests and interpretations* calls for equivalence to be established between the computerized and original versions of the assessments. This necessitates comparisons of means, distributions, ranking of scores and correlations with other variables. Tseng *et al* (1998) argue that for equivalence to be truly established, individual characteristics should not differentially affect a person's responses to a particular administration mode of an assessment.
>
> Brosnan in Brown, Race and Bull, 1999, p49

To be widely used, the web-based administration of these instruments must be characterized in terms of reliability, and results from the web-based administration of these instruments must be statistically compared to results from standard paper administration. If measurements from web-based administrations are explored, they can be corrected or calibrated to paper-based administrations. Therefore, the purpose of this study is to begin this process by examining the differences in paper-based and web-based administrations of the Force Concept Inventory.

Method

Data Source/Participants

The participants made up a sample of 1313 students, 233 (19.90%) women and 938 (80.10%) men. The majority of the students were Caucasian, in the age range of 18 to 22 years and therefore and age and ethnicity were not considered further. The participants

were all were students from an introductory physics courses taught at medium sized university in the midwest during the Spring of 2000.

Instruments

The Force Concept Inventory (FCI) is a 30 item multiple choice test which "requires a forced choice between Newtonian concepts and common-sense alternatives" (Hestenes, Wells, & Swackhamer, 1992, p. 142). The filler task was a 34 item Likert instrument, the Values and Attitudes about Science Survey (VASS).

Procedure

During the Spring of 2000, two introductory physics classes participated. Each class was divided into two roughly equal (verified within five percent for each recitation and for the class overall) half-class groups by assigned all enrolled students to two half-class groups by the random criteria of whether their eight digit student identification number ended in an even or an odd digit. During the first week of the semester, thirty minutes was devoted to testing. In each class, one half-class group was administered a paper-based FCI and then asked to complete a web-based filler task (VASS) in the next seven days. The other half-class group was administered a paper-based filler task (VASS) and then asked to complete the web-based FCI in the next seven days. The filler task was a questionnaire about student's attitudes towards science (VASS). This entire data collection process was repeated during the last week of the semester with students who started the semester taking a paper-based FCI ending the semester taking a web-based FCI.

Each student was supplied with the web address for the test appropriate to their assigned half-class group. No training was provided to the students for taking either test on the web. Further, there was no attempt to authenticate the web users. Rather, each student's work was accepted as their own. Times for overall test completion were recorded along with the time and date the student submitted the test form for grading. This information was used to ensure that students took no longer than 30 minutes to complete the test and

that they took the test within the seven day period.

All of the tests were graded as to completeness and counted as the as equivalent of one homework or quiz depending upon the class. Grades of 0, 1, or 2 of two points possible were assigned for satisfactory completion of the paper-based and web-based FCI and VASS tests. With respect to final class grades, students' participation comprised four points out of one thousand total points, so that completion or non-completion had negligible impact.

Results

As a result of the paper-based and web-based administrations, 1313 students participated in the study. Pre-test data collected at the beginning of the semester totalled 1173 usable tests while the post-test data collected during the last week of the semester totalled 825 usable tests. (Tests that were turned in after the seven day period or that were taken for longer than 30 minutes were deemed unusable for the purpose of this analysis.) Student scores on the FCI were calculated by adding the total number of correct answers with a total possible FCI score being 30. The pre-test mean was 15.25 ($\underline{N}$ = 1173, $\underline{SD}$ = 5.69) and the post-test mean was 19.17 ($\underline{N}$ = 825, $\underline{SD}$ = 6.44).

Paper-based Versus Web-based FCI Student Scores

Previous research has indicated differences in FCI scores due to gender. Therefore, to examine differences in paper-based and web-based FCI student scores a 2 X 2 ANOVA was used (2 genders, 2 types of FCI administration). An alpha level of .01 was used for all statistical tests. For both the pre- and post-tests, significant differences were found for the main effect gender and no significant differences were found for the main effect, type of FCI administration. For the first-order interactions, no significant differences were found due to type of FCI administration (see Table 1 for statistics).

Table 1

Two-Way ANOVA summary table for gender, type of FCI administration for FCI Pre-Test.

Source	df	MSe	F
Pre-Test (n=1173)			
gender	1	3285.11	111.31*
administration	1	1.01	.03
gender x administration	1	19.02	.64
Post-Test (n=825)			
gender	1	2345.73	60.44*
administration	1	24.22	.62
gender x administration	1	19.07	.49

*$p < .01$

To further examine potential differences in the student scores on the paper-based and web-based administrations of the FCI, Cronbach's alpha was calculated separately for the paper-based and web-based administration of the FCI for the pre-test, post-test (see Table 2). As we see from the table, these alpha levels appear to be comparable.

Table 2

Cronbach's Alpha for paper-based and web-based versions of FCI pre-test, post-test.

	pre-test		post-test	
Version	n	alpha	n	alpha
Paper	614	.83	407	.87
Web	559	.84	418	.89

Paper-Based Versus Web-based Individual FCI Items

Differences in the paper-based and web-based administrations of the FCI for individual items was explored using t Tests. A probability level of .01 was used for all statistical tests. The F statistic was used to determine whether the variances of the paper-based and web-based administrations of each item were equal. Only one item (number 6) demonstrated a significant difference between paper-based and web-based administrations and this occurred only during the post-test (see Table 3 for statistics).

Table 3

Results of t Tests* for paper-based and web-based administrations of FCI items for pre- and post-test times

	Pre-test	Post-test		Pre-test	Post-test
Item	F, prob<F	F, prob<F	Item	F, prob<F	F, prob<F
Item 1	1.00, .98	1.09, .38	Item 16	1.03, .74	1.27, .02
Item 2	1.01, .90	1.04, .71	Item 17	1.09, .30	1.09, .36
Item 3	1.03, .73	1.10, .33	Item 18	1.01, .86	1.00, .99
Item 4	1.04, .61	1.02, .86	Item 19	1.00, .98	1.02, .85
Item 5	1.18, .05	1.00, .99	Item 20	1.02, .81	1.00, .96
Item 6	1.17, .06	1.46, .001*	Item 21	1.01, .93	1.00, .97
Item 7	1.02, .80	1.09, .38	Item 22	1.00, .98	1.10, .31
Item 8	1.08, .36	1.10, .35	Item 23	1.01, .89	1.00, .99
Item 9	1.01, .91	1.01, .89	Item 24	1.21, .02	1.13, .21
Item 10	1.02, .83	1.00, .97	Item 25	1.04, .63	1.01, .91
Item 11	1.00, .97	1.01, .94	Item 26	1.01, .87	1.02, .83
Item 12	1.02, .81	1.16, .13	Item 27	1.13, .15	1.01, .88
Item 13	1.05, .58	1.05, .61	Item 28	1.00, .99	1.00, .98
Item 14	1.04, .64	1.13, .17	Item 29	1.00, .99	1.02, .86
Item 15	1.01, .93	1.00, .98	Item 30	1.02, .77	1.02, .87

df = (614, 569) for all pre-tests, and df = (418, 407) for all post-tests

Chi Square tests of the paper-based and web-based administrations of each item were conducted to determine whether the response patterns (patterns of A, B, C, D, or E responses) of the paper-based and web-based administrations differed. A probability level of .01 was used for all statistical tests. Two items demonstrated a significant difference in the response patterns for paper-based and web-based administrations at both pre- and

post-test (numbers 17 and 30, see Table 4).

Table 4

Results of χ^2 tests for paper-based and web-based administrations of FCI items for pre- and post-test.

	Pre-Tests	Post-Tests		Pre-Tests	Post-Tests
Item	χ^2, p	χ^2, p	Item	χ^2, p	χ^2, p
Item 1	1.73, .78	2.15, .71	Item 16	5.26, .26	6.95, .13
Item 2	6.29, .18	8.02, .09	Item 17	563.36, .001*	272.10, .001*
Item 3	3.94, .41	6.88, .14	Item 18	1.01, .91	3.89, .42
Item 4	6.15, .19	9.62, .05	Item 19	0.99, .91	9.53, .05
Item 5	7.67, .10	3.72, .45	Item 20	3.69, .45	1.17, .88
Item 6	11.68, .02	11.63, .02	Item 21	11.26, .02	4.26, .37
Item 7	8.49, .08	5.61, .23	Item 22	3.08, .09	6.13, .19
Item 8	10.41, .03	8.52, .07	Item 23	6.92, .14	1.56, .82
Item 9	4.27, .37	.43, .98	Item 24	6.42, .17	5.40, .25
Item 10	3.91, .42	.36, .99	Item 25	10.04, .04	4.50, .34
Item 11	4.60, .33	4.43, .35	Item 26	2.54, .64	10.17, .07
Item 12	5.32, .26	2.21, .70	Item 27	6.64, .16	6.80, .15
Item 13	12.09, .02	4.10, .39	Item 28	5.55, .24	.37, .98
Item 14	4.36, .36	7.48, .11	Item 29	4.76, .31	6.71, .15
Item 15	2.01, .73	5.74, .22	Item 30	14.74, .01*	14.75, .01*

df = 4 for all tests

Summary of Results

The results of these analyses demonstrated little appreciable difference on FCI scores or items based on the type of administration. While the results of a 2 way ANOVA demonstrated differences in FCI student scores due to gender and time of administration,

none of these differences was influenced by the type of test administration. Additionally, FCI student scores were comparable with respect to reliability. For individual FCI items, paper-based and web-based comparisons were made by examining potential differences in item means and by examining potential differences in response patterns. Again, very few differences in item means (as demonstrated by t Tests) and in response patterns (as demonstrated by Chi Squares) were found between the paper-based and web-based FCI items. In summary, the web-based administration of the Force Concept Inventory appears to be as efficacious as the paper-based application.

Significance and Discussion

This study sought to examine potential differences in paper-based and web-based administrations of the Force Concept Inventory. The results of these analyses demonstrated no appreciable differences on FCI scores or items based on the type of administration. While the results of a 4 way ANOVA did demonstrate differences in FCI student scores due to different sections, courses, and gender, *none of these differences were influenced by the type of test administration.* FCI student scores were comparable with respect to both reliability and predictive validity. For individual FCI items, paper- and web-based comparisons were made by examining potential differences in item means and by examining potential differences in response patterns. *Again, no differences in item means (as demonstrated by t Tests) and no differences in response patterns (as demonstrated by Chi Squares) were found.* In summary, the web-based administration of the Force Concept Inventory appears to be as efficacious as the paper-based administration.

Although this study reports no differences between web and paper-administrations of the FCI, there are a number of issues related to web-administered testing of concern to students, instructors and researchers. The first of these is academic dishonesty. In our study, students were awarded only a small grade (1-3 points maximum from 1000 total for the course) for completing the survey. We wanted to encourage students to participate and to

be conscientious in their responses, yet minimize the incentive to cheat. We did not prevent students from copying or printing out the test, nor did we authenticate that the students were who they claimed to be. There is no practical way of doing these things without requiring students to take the test in a proctored computer lab; a solution which has been used at other institutions (e.g. Harvard). In earlier research, we developed the expertise to reduce the likelihood of inappropriate printing or sharing of the instrument by restricting access to the online tests with a changing login and password that was only functional for limited times at the start and end of the semester. Originally, our software reported the number of correct responses for the instrument back to the student; we removed this feedback after having an experience where a student repeatedly submitted the survey while varying answers trying to maximize their score. Now the instrument simply thanks the student upon submission.

Another issue related to web-administered tests is the resolution of the student's computer video monitor. Computer video monitors have a much lower resolution than paper printouts (typically 72 dots per inch vs. 600 dots per inch). In the present study, the paper-administered FCI was a direct printout of the web pages (Fig 1). However, the finer resolution of the laser printer made it easier to read both the text and graphics, particularly the vectors and dotted lines which indicated trajectories. While Clausing and Schmitt (1989, 1990a, 1990b) found that with reasonable diligence, there was no a difference in reading errors between computer video monitors and paper-printed tests, the finer paper resolution may still be more comfortable to work with.

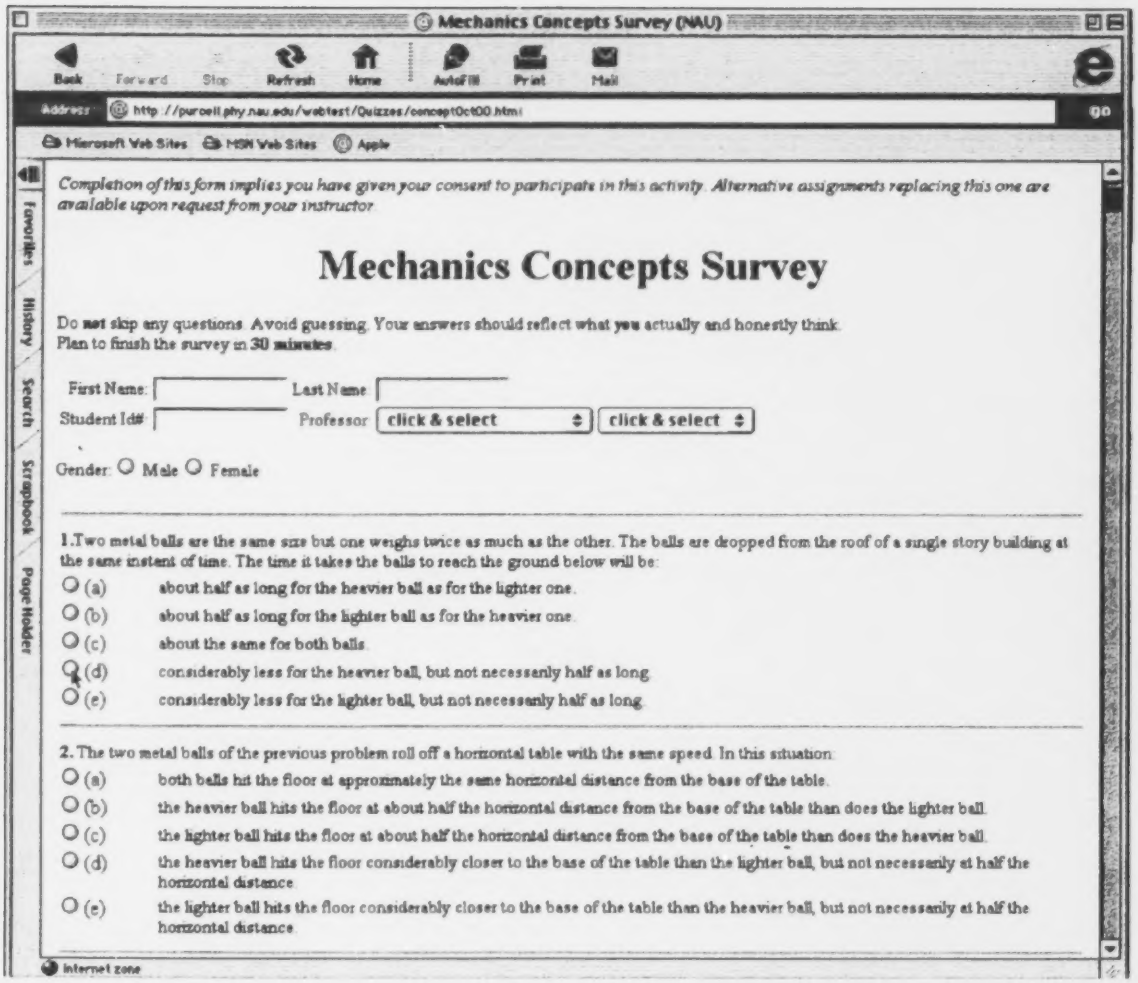
Mechanics Concepts Survey (NAU)
Back
Forward
Stop
Refresh
Home
AutoFill
Print
Mail
Address
http://purcell.phy.nau.edu/webtest/Quizzes/concept0ct00.html
go
Microsoft Web Sites
MSN Web Sites
Apple
Favorites
History
Search
Scrapbook
Page Holder
Completion of this form implies you have given your consent to participate in this activity. Alternative assignments replacing this one are available upon request from your instructor.
Mechanics Concepts Survey
Do not skip any questions. Avoid guessing. Your answers should reflect what you actually and honestly think.
Plan to finish the survey in 30 minutes.
First Name:
Last Name:
Student Id#:
Professor
click & select
click & select
Gender: Male Female
1. Two metal balls are the same size but one weighs twice as much as the other. The balls are dropped from the roof of a single story building at the same instant of time. The time it takes the balls to reach the ground below will be:
(a) about half as long for the heavier ball as for the lighter one.
(b) about half as long for the lighter ball as for the heavier one.
(c) about the same for both balls.
(d) considerably less for the heavier ball, but not necessarily half as long.
(e) considerably less for the lighter ball, but not necessarily half as long.
2. The two metal balls of the previous problem roll off a horizontal table with the same speed. In this situation:
(a) both balls hit the floor at approximately the same horizontal distance from the base of the table.
(b) the heavier ball hits the floor at about half the horizontal distance from the base of the table than does the lighter ball.
(c) the lighter ball hits the floor at about half the horizontal distance from the base of the table than does the heavier ball.
(d) the heavier ball hits the floor considerably closer to the base of the table than the lighter ball, but not necessarily at half the horizontal distance.
(e) the lighter ball hits the floor considerably closer to the base of the table than the heavier ball, but not necessarily at half the horizontal distance.
Internet zone

Figure 1: The FCI in scrolling format, matched to standard paper instrument.

In addition, it was difficult for students using a smaller computer monitor to see several test questions together with the accompanying diagrams. Conversely, printed pages afford students the opportunity to easily flip back and forth or lay successive pages side by side. For the web-administrations, this can only be accomplished by the unwieldy process of scrolling back and forth. A new version of our software for administering instruments works around this by allowing flipping back-and-forth style access to other items on the instrument while simultaneously collecting latency data by the individual item (see Fig 2).

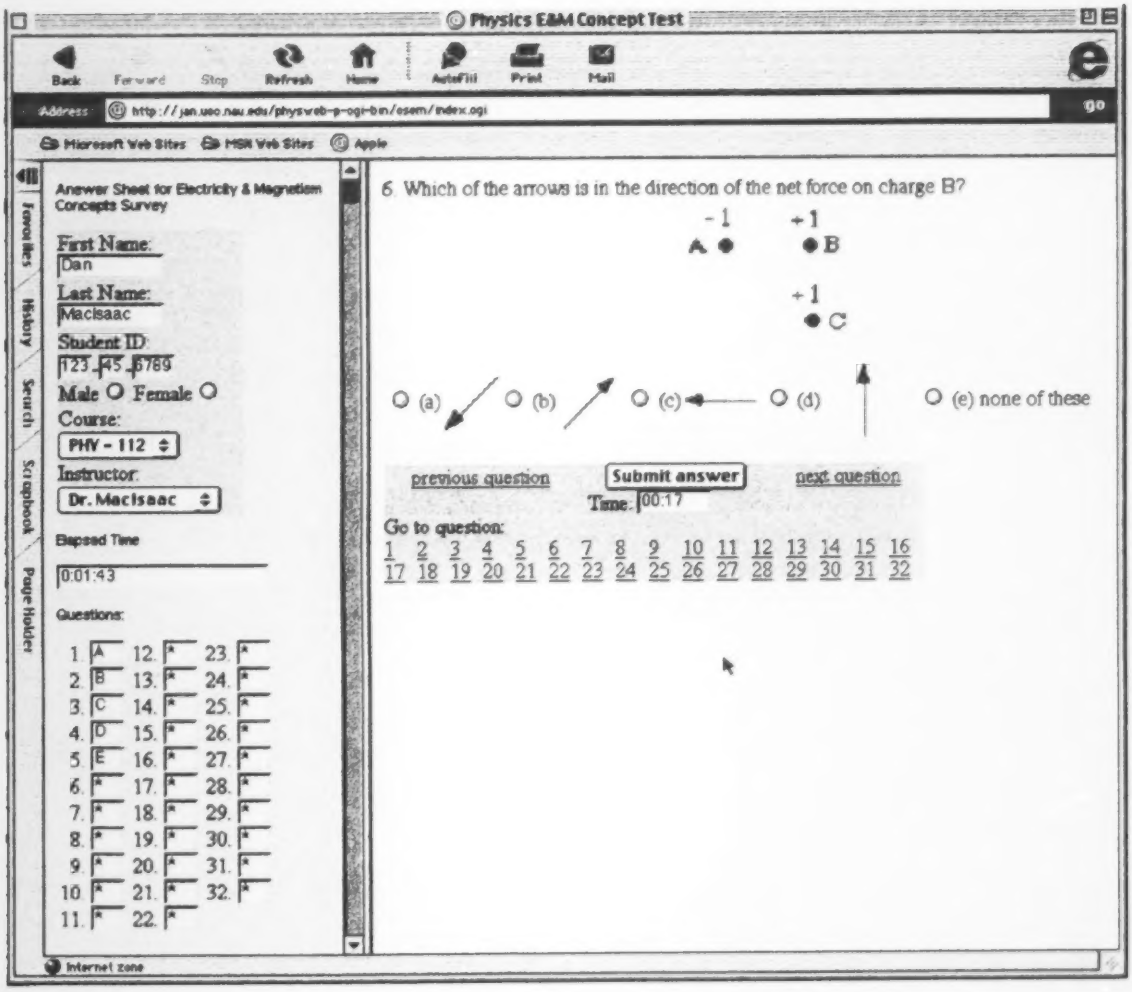

Figure 2: A web-administered standard instrument in item latency format.

Finally, the paper-administered FCI coding sheets demonstrated problems. In our study, the optically-encoded scanned bubble sheets produced errors due to skipped rows of questions and incomplete erasures. We eliminated such errors from our data set by rigorously proofreading and screening bubble sheets prior to scanning, and by comparing scanner output files to the original bubble sheets. Such proofing is unlikely to occur with typical paper-administrations, as it poses a significant additional burden on the instructor. Eliminating the use of bubble sheets and allowing students to mark directly on the test might

BEST COPY AVAILABLE

alleviate this problem, but would complicate the grading process. In comparison, the web administered FCI used "radio buttons" for responses. These buttons accurately code only one solution per question, allowed students to cleanly change responses (i.e. no erasing), and aligned each and every response with the question text and graphics on the screen.

Conclusions and Implications

This study demonstrated no differences between the paper-based and web-based administration of a major standardized physics test, the Force Concept Inventory. The main implication of this finding is that, at least for the FCI, web-based administrations could be used in place of paper- administrations, thus saving precious instructional time, reducing the administrative overhead associated with testing, grading, and photocopying thus cutting the costs associated with large scale data collection. Further, web-based administrations offer information that paper-based administrations do not. For example, item latency and completion data can be collected.

We are extending this research by investigating the possibility of creating a web-based "Physics Testing Center" that could administer tests and feed resulting measurements directly into a modern database. Such a testing center would allow for the routine collection of conceptual and attitudinal data and be available for longitudinal studies of student learning and instruction. This would enhance our understanding of programs and pedagogy both inside and outside our university. Another use of a Physics Testing Center would be the opportunity for researchers to pilot and standardize new instruments by providing access to large numbers of student participants. Faculty from other departments have seen our efforts and have started the design and develop of 'screening' instruments intended for student guidance and placement in the gatekeeper science courses at NAU.

Along these lines, the authors have begun to collaborate with other researchers and institutions in an attempt to create such a centralized web-based testing center and common database. In addition, we are expanding our on-line standardized testing effort to include other instruments. Specifically, we are readying the Conceptual Survey in Electricity and

Magnetism (Hieggelke, Maloney, O'Kuma, & van Heuvelen, 1996) for web-based administration.

Author Notes

Financial Support:

This research was partially supported by Northern Arizona University Organized Research Grant funds and by the NAU Department of Physics & Astronomy. Additional support was provided by National Science Foundation funding through the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) and the US Department of Education funded AriZona Teacher's Excellence Coalition (AzTEC). Professors David Elmore, Andrew Hirsch and Edward Shibata of the Purdue University Department of Physics invited and supported this study.

Acknowledgments:

The authors wish to acknowledge the helpful contributions of Nate Davis, Brian Nance and Eric Tse who assisted with HTML and data coding and were funded by an NAU Hooper Undergraduate Research Fellowship. Valuable comments and suggestions regarding the statistical comparisons of item response patterns was provided by Professor Philip Sadler of Harvard University.

References

Arons, A.B. (1984). Computer-based instructional dialogs in science courses. Science, 224, 1051.

Arons, A.B. (1986). Overcoming conceptual difficulties in physical science through computer-based Socratic dialogs. In A. Bork & H. Weinstock (Eds.), Designing computer-based learning materials: NATO ASI Series Vol F23. Berlin: Springer-Verlag.

Bork, A. (1981). Learning with computers. Bedford MA: Digital Press.

Brosnan, M. (1999). Computer anxiety in students: Should computer-based assessment be used at all? In Brown, S., Race, P. & Bull, J. (1999) (Eds), Computer-assisted assessment in higher education. London: Kogan-Page.

Cann, A.J. & Pawley, E.L. (1999). Automated online tutorials: New formats for assessment on the WWW. In Brown, S., Race, P. & Bull, J. (1999) (Eds), Computer-assisted assessment in higher education. London: Kogan-Page.

Champagne, A.B., Klopfer, L.E., & Anderson, J.H. (1980). Factors influencing the learning of classical physics. American Journal of Physics, 48, 1074-1079.

Clausing, C.S., & Schmitt, D. R. (1989). The effects of computer usage on computer screen reading rate. Paper presented at the annual meeting of the Mid-South Educational Research Association. Little Rock, AK.

Clausing, C.S., & Schmitt, D. R. (1990a). Paper versus CRT: Are reading rate and comprehension affected? Proceedings of Selected Paper Presentations at the Convention of the Association for Educational Communications and Technology.

Clausing, C.S., & Schmitt, D. R. (1990b). Does computer screen color affect reading rate? Paper presented at the annual meeting of the Mid-South Educational Research Association. New Orleans, LA.

Danson, M. (1999). Optimal mark reading as CAA: Experiences at Loughborough University. In Brown, S., Race, P. & Bull, J. (1999) (Eds), Computer-assisted assessment in higher education. London: Kogan-Page.

Hake, R.R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 64-74.

Hallouin, H. (1998). Views about science and physics achievement: The VASS story. Proceedings of the Third International Conference on Undergraduate Physics Education. College Park, MD.

Hallouin, H., & Hestenes, D. (1996). The initial knowledge state of college physics students. American Journal of Physics, 53, 1043-1055.

Association. New Orleans, LA.

Harvey, J. & Mogey, N. (1999). Pragmatic issues when integrating technology into the assessment of students. In Brown, S., Race, P. & Bull, J. (1999) (Eds), Computer-assisted assessment in higher education. London: Kogan-Page.

Hestenes D., & Hallouin, I. (1995). Interpreting the Force Concept Inventory: A response to March 1995 critiques by Huffman and Heller. The Physics Teacher, 33, 502.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. The Physics Teacher, 30, 141-158.

Hieggelke, C.J., Maloney, D., O'Kuma, T., & Van Heuvelen, A. (1996). Electric and magnetic concept inventory. AAPT Announcer, 26, 80.

Huffman D., & Heller, P. (1995). What does the Force Concept Inventory actually measure? The Physics Teacher, 33, 138.

Linn, R.L. (Ed.). (1978) Educational Measurement (3rd ed.). New York: Macmillan.

MacIsaac, D.L. (1999a). Department of physics and astronomy assessment plan and preliminary data report. Unpublished manuscript available from the author.

MacIsaac, D.L. (1999b). Comparing web and paper-based administration of standardized introductory physics conceptual and attitudinal instruments. Final Program and Abstracts: 72nd NARST Annual Meeting 1999, 204.

Redish, E.F., Saul, J.M., & Steinberg, R.N. (1998). Students expectations in introductory physics. American Journal of Physics, 6, 212-224.

Straetmans, G., & Eggen, T. (1998). Computerized adaptive testing: What it is and how it works. Educational Technology, JanFeb98, 45.

Titus, A.P., Martin, L.W., & Beichner, R.J. (1998). Web-based testing in physics education: Methods and opportunities. Computers in Physics, 12, 117.

Waugh, M.L. (1985). Effects of microcomputer-administered diagnostic testing on immediate and continuing science achievement and attitudes. Journal of Research in Science Teaching, 23, 793.

Wise, S.L., & Plake, B.S. (1990). Computer-based testing in higher education. Measurement and Evaluation in Counseling and Development, 23, 3.

Zandvliet, D., & Farrager, P. (1997). A comparison of computer-administered and written tests. Journal of Research on Computing in Education, 29, 423.